Exclusion or Efficiency: Understanding Perspectives about AI Ethics Among Charity Workers in the United Kingdom

SAKINA HANSEN, London School of Economics, United Kingdom

The widespread use of AI tools across society has impacted many different individuals, organizations and stakeholders. The ethical issues that arise are of great focus of academic research, but there is significantly less engagement with the different types of organizations effected, particularly how the charity sector is effected. I conducted a pilot empirical study consisting of semi-structured qualitative interviews with three employees that work in some capacity with data or technology at a charity, from three different charities that operate in the United Kingdom. This work offers insight into the unique challenges and perspectives faced by charities. I found that they view the ethical risk of AI primarily through its ability to be an exclusionary tool, and view the positives of AI primarily through its ability to be an efficiency tool.

Keywords: AI ethics, charity, qualitative research, interviews, thematic analysis

Reference Format:
Sakina Hansen. 2025. Exclusion or Efficiency: Understanding Perspectives about AI Ethics Among Charity Workers in the United Kingdom. In *Proceedings of Fourth European Workshop on Algorithmic Fairness (EWAF'25).* Proceedings of Machine Learning Research, 17 pages.

1 Introduction

Artificial intelligence (AI) is increasingly deployed in areas that have significant impacts on society and individuals. A number of ethical risks are associated with AI such as discriminatory algorithms [11, 12], risks of a decrease in autonomy as it becomes more difficult to challenge decisions [22, 23], and decreasing trust caused by a lack of transparency which can potentially violate individual rights [21]. This has propelled AI ethics research directives such as explainability and interpretability [1, 18], fairness [5, 13, 17], and algorithmic auditing [4, 6, 20], among others. Much of these proposed solutions come from academic and corporate research, where the researchers don't directly deal with the day-to-day effects of algorithmic-decision making. The hope is that explainability, fairness and audit methods are practically useful to those who deal with algorithms in practice and who wish to understand, evaluate or audit algorithms. Those who do deal with the day-to-day effects are referred to as 'stakeholders', which can include companies, government, charities, individuals and others. However, whether these methods help with the challenges faced by stakeholders is not well studied, particularly for charities.

While companies often have the means to use and assess algorithms, individuals affected by decisions or charities most likely do not due to lack of resources. Therefore, the discussion about these stakeholders in the academic literature is often spoken of in the abstract, with a lack of engagement with their perspectives. Browne et al. [7] conducted a qualitative study on tech worker perspectives on ethical issues in AI development using interviews, but

Author's Contact Information: Sakina Hansen, London School of Economics, London, United Kingdom, s.a.hansen1@lse.ac.uk.

EWAF'25, June 30–July 02, 2025, Eindhoven, NL

this was focused on internal company development and not on users or people effected by the algorithms. Charities are likely to face unique challenges and opportunities when it comes to AI. Charity work does essential work across the country, so it is necessary to understand how technological advances affect them. How charity responds will have knock-on effects for the rest of society, making this a very important perspective to understand. Additionally, understanding their perspectives is informative for technical work on AI ethics as well as for AI policy development. Therefore, this research aims to answer the following question: ***What are the perspectives and experiences UK charity workers have with AI and AI ethics?***

2 Methodology

To understand the perspectives of charity workers in the UK, I collected data by conducting interviews with three individuals who work for charities based in the UK. Interviews are effective at generating a broad range of topics related to a research question [10], so they provide a better overview of the themes related to the research question. All the interviewees work for a charity in the UK and are directly working on responses or strategies regarding digital services, data or AI. This decision was made in order to select interviewees who would be particularly informed on how their charity, or the charity sector in general, views AI. Two interviewees worked at charities that offer advice to other charities on how to use data and technology to improve their services, while the third interviewee was part of a data science team at a charity that provides free advice on a range of issues, from debt to immigration, across the UK. The interviewees were recruited through my personal network and through a Slack channel for charities that use data science.

I want to understand how the interviewees' experiences and opinions on AI are informed by their work in charities. For this reason I took a relational and semi-structured approach to the interviews, as I want "to learn how they make sense of the world by engaging them in dialogue" [9]. Relational interviewing is effective at understanding how people explain processes of change, which is what we're seeing as AI is creating a process of change in many aspects of society. Semi-structured interviews were chosen to to allow for flexibility for different topic trajectories [16]. An interview topic guide was developed (see Appendix C) to guide the interviews.

The analysis method chosen for analyzing the interview data was thematic analysis, primarily following the six-steps proposed by [3]. This method allows for understanding the many different ways people may understand a topic or belief, rather than aiming to find a single answer. The research question is about understandings and so is suited to thematic analysis. The six-step method consists of coding the material, identifying themes, constructing thematic networks, describing and exploring thematic networks, and summarizing thematic networks. The process is iterative, with reading repetitions to rigorously understand how the different basic themes are communicated. This also helps to effectively draw connections between themes themselves and between themes and contextual factors. There are eleven basic themes identified that are relevant to the question, which were then grouped into four organizing themes which clarify the connections between the basic themes. These organizing themes were than classified into two global themes. The codebook and thematic network produced from this analysis can be found in Appendix B and Appendix A, respectively.

3 Findings

Thematic analysis on the interview data revealed two global themes that address the research question, namely that AI is understood as either an exclusionary tool or an efficiency tool. These themes both show how charity work informs the interviewees' opinions on the challenges and possibilities of AI. The full thematic network is available in Appendix A, and the codebook is available in Appendix B.

3.1 AI as an Exclusionary Tool

Overall, charity workers understand the ethical issues of AI through a lens of exclusion, first due to how AI can exclude charities from society due to AI development ties to capitalist incentives, and second due to AI excluding individuals and charity clients through marginalization.

The former is understood to be because of AI being developed primarily in the service of business and government, without considering knock-on effects to charity: *"The thing that leaves me still unconvinced is how tied the growth of AI is to extractive business models."* (Interviewee 1). This leads to a feeling of exclusion for charities, expressed in terms of equality across society: *"I most align with ... community driven governance, where the impacts of AI are equally shared across society rather than driven specifically by a business model"* (Interviewee 1). The feeling that technological development and government exclude people as well as exclude the specific needs of charities is illustrated by the following quote: *"A chatbot would not be suitable for 20% of our clients... The risk there is for that like cohort of people they will become more excluded. These pressures come from government funders looking to or making assumptions about how services can be more cost effective without thinking about how they can be more inclusive."* (Interviewee 3)

This understanding of AI's relationship to capitalist development is shown in fears surrounding monopolization of the charity sector which would exclude smaller charities which do essential work: *"[There's a] risk in the charity sector, not intentional, where the very large charities take over as digital begins to accelerate. They have very large, capable teams and they're comfortable around tech, they're able to invest in it. If they then become super dominant, it will cut the throat of other charities."* (Interviewee 2). This fear is also driven by lack of funding in technical skills for the charity sector: *"One of their main fears about AI is predatory competitors who make better use of these technologies... there's a lot of fear about other organizations having like a technical competitive advantage."* (Interviewee 3).

The understanding that AI excludes the public is made sense of through how AI is uniquely able to marginalize individuals. The interviewee's all work at charities that deal with vulnerable, already marginalized people, and so this charity work experience is reflected in how they see AI further marginalizing people. I first focus on the theme related to what one interviewee termed the 'super-excluded'. This term refers to how those people in society who are digitally excluded will become further excluded as AI advances, and reducing the ability of charities to help them: *"The super-excluded ... I think those people are at risk of just falling off the edge. So there'll be a much smaller group, but they will be much more excluded."* (Interviewee 2); *"I think there's just a big risk of leaving a lot of people behind. And also if, essentially delivering services that are worse and don't have that kind of human element to them."* (Interviewee 3).

More familiar to the AI ethics academic field, is the theme that AI can exclude and marginalize people through predictions. It is understood that this could come from charities themselves and therefore people won't get the

essential help they need: *"we're very conscious that the risk due to underlying bias ... is that people who are already kind of disadvantaged, so minoritized clients, will be disadvantaged even more."* (Interviewee 3), *"It may be our algorithms that are making the inappropriate decisions if we're not careful"* (Interviewee 2).

There is also a significant worry about not being able to respond to the decisions of external algorithms. Many resources are already difficult enough for some clients to get access with algorithms only making it worse: *"... we need algorithmic transparency, because without it well... as I understand it, it's already challenging enough to navigate [the Personal Independence Payment process] and why decisions have been made"* (Interviewee 3); *"... they can have about 100% turnover rate*[1] *because they can challenge the decision by understanding how those decisions have been made. They have very low digital maturity, they can't understand decisions that are made algorithmically. In that kind of predictive future, you see a 100% turnover rate get down to 0."* (Interviewee 1); *"[we saw] clients of color were being charged more for their vehicle insurance as a result of algorithmic decision making ... I think we're going to be seeing this increasingly in the coming years."* (Interviewee 3).

This connects to a very relevant theme that the AI ethics field in academia should consider; algorithmic decision making is causing charities to have to wait for groups of people to be discriminated against before they are able to detect the pattern of discrimination. This allows for exclusion of individuals because by the time the pattern emerges, a negative decision has already had impact on people's lives: *"At the moment we haven't got the mechanism to say .. . to quickly correct.... If the person who is like delivering the decision doesn't understand how its being made, or because there's so many inputs, no one really understands how it's been made. Then the time lag from decision through to consequence... that's a real kind of challenging thing."* (Interviewee 1). This is two-fold, in that it excludes the clients from resources, but also excludes charities: *"You've got charities who are trying to campaign and try to influence change, not being able to, so investigative journalists and data scientists being able to take on that role ... It shifts from supporting individuals who have had a negative decision, to supporting cohorts when there's enough data to spot patterns..."* (Interviewee 1).

Opinions on AI Ethics Solutions. A topic of focus of the interviews was about finding out how appropriate some of technical ethics AI solutions are. The interviewees expressed interest in presented AI ethics methods such as fairness and explainability (see Appendix D for methods that were described to the interviewees), but only for basic themes of AI bias, and mostly for internal AI bias, as it would be more difficult to apply these methods to external algorithms. There are many other ethical issues highlighted by the global theme of exclusion that AI ethics solutions are not yet addressing completely (and maybe are unable to), such as super-exclusion and lack of individual support. These areas would be promising avenues for future AI ethics research.

3.2 AI as an Efficiency Tool

The second global theme that emerged was that the role of AI in charity can be understood in terms of its ability to make charities more efficient by helping charities have more efficient finances and more efficient services, a similar finding to [19]. There was real optimism expressed that AI could significantly reduce financial burden: *"The sector is in desperate straits. It never recovered from 2008. There's COVID. There's cost of living crisis ... There's a huge funding gap, there are increasing numbers of charities going to the wall ... AI will do two things. It will augment*

[1]Turnover rate here refers to a charity's ability to help a client overturn a negative decision, for example a rejection of a benefit claim.

your capabilities, and it will reduce admin workload ... It can't fix the funding gap, but it can depressurize it." (Interviewee 2). This also extended to reducing financial burden on workers so that their lives could improve: *"The benefit of those efficiencies is that they have more time to spend with their friends, families and communities ... we can benefit from all these efficiencies as a society, rather than just those who are able to create businesses and make more efficient businesses"* (Interviewee 1). There is a caveat in this quote, which was consistent among the interviewees, that these possible positives that AI could bring, are dependent on fixing the problems of exclusion highlighted by the previous global theme.

AI is understood to also be able to improve efficiencies in content generation which would have usually come at a financial cost: ... *the future of generating content. So we employ, you know, content designers to essentially interpret the law. So the law is not designed to be user friendly as it were ... But one of the areas of change is the ability for generative AI to interpret the law into accessible text"*

In terms of more efficient services, it is believed that AI could make charity services faster: *"There are loads of positives around efficiencies, and enabling people to get immediate decisions."* (Interviewee 1) and *"From a generative AI point of view, we've started work to see how we can make use of these technologies in improving the provision of advice... we have more demand for advice than we can supply. So it's important to us to see how we can introduce efficiencies in that workflow."*

Additionally, another way it is understood that AI can make services more efficient is by improving accessibility, if AI is developed and deployed in a way that considers how it will effect marginalized people: *"I also think there's an opportunity if you approach it in the right way to make services more inclusive. ... Using AI, but with an intermediary who works with a client who would otherwise be excluded from its benefits. And I think also, there's the opportunity through challenging and mitigating bias to make outcomes more equitable as well."* (Interviewee 3). Again, there is a caveat outlined here that this success of efficiency depends upon addressing the problems of exclusion.

4 Conclusion

The charity sector faces unique challenges when it comes to AI and AI ethics. In this work we conducted interviews with three individuals who work for charities in the UK, in roles that involve dealing with digital services, data or AI. We found that data focused charity workers understand AI primarily as an exclusionary tool or an efficiency tool. The interviewees were all fairly senior in their roles, and all worked with data. This means that while they have a well-informed opinions that are helpful to answer this research question, they represent a small sample compared to charity workers in general. Further interviews or focus groups that have participants with different job roles, from varying organization sizes and domains, could improve understanding of possible heterogeneity in views of AI ethics. For this reason, this research has not yet reached saturation. We hope this work encourages more engagement and qualitative research to understand the specificity of AI related issues that are faced by different people and organizations in society and how this should shape AI development and AI policy.

Acknowledgments

This work received ethics approval from the London School of Economics. I would like to thank the interviewees for taking part, Raphael Susewind for his great teaching on qualitative methods, and to the reviewers for their helpful comments.

References

[1] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

[2] Daniel W Apley and Jingyu Zhu. 2020. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82, 4 (2020), 1059–1086. https://doi.org/10.1111/rssb.12377

[3] Jennifer Attride-Stirling. 2001. Thematic networks: an analytic tool for qualitative research. *Qualitative research* 1, 3 (2001), 385–405. https://doi.org/10.1177/146879410100100307

[4] Jack Bandy. 2021. Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 74 (April 2021), 34 pages. https://doi.org/10.1145/3449148

[5] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2023. *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press.

[6] Abeba Birhane, Ryan Steed, Victor Ojewale, Briana Vecchione, and Inioluwa Deborah Raji. 2024. AI auditing: The Broken Bus on the Road to AI Accountability. In *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. 612–643. https://doi.org/10.1109/SaTML59370.2024.00037

[7] Jude Browne, Eleanor Drage, and Kerry McInerney. 2024. Tech workers' perspectives on ethical issues in AI development: Foregrounding feminist approaches. *Big Data & Society* 11, 1 (2024). https://doi.org/10.1177/20539517231221780

[8] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232. https://doi.org/10.1214/aos/1013203451

[9] Lee Ann Fujii. 2017. *Interviewing in social science research: A relational approach*. Routledge. https://doi.org/10.4324/9780203756065

[10] Greg Guest, Emily Namey, Jamilah Taylor, Natalie Eley, and Kevin McKenna. 2017. Comparing focus groups and individual interviews: findings from a randomized study. *International Journal of Social Research Methodology* 20, 6 (2017), 693–708. https://doi.org/10.1080/13645579.2017.1281601

[11] Pauline T Kim. 2017. Auditing Algorithms for Discrimination. *U. Pa. L. Rev. Online* 166 (2017), 189.

[12] Nima Kordzadeh and Maryam Ghasemaghaei. 2022. Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems* 31, 3 (2022), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

[13] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual Fairness. *Advances in Neural Information Processing Systems* 30 (2017).

[14] Joshua Loftus, Lucius Bynum, and Sakina Hansen. 2024. Causal Dependence Plots. *Advances in Neural Information Processing Systems* 37 (2024), 112656–112683.

[15] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 4768–4777.

[16] Danielle Magaldi and Matthew Berler. 2020. Semi-structured Interviews. *Encyclopedia of Personality and Individual Differences* (2020), 4825–4830. https://doi.org/10.1007/978-3-319-24612-3_857

[17] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6, Article 115 (July 2021), 35 pages. https://doi.org/10.1145/3457607

[18] Christoph Molnar. 2022. *Interpretable Machine Learning* (2 ed.). https://christophm.github.io/interpretable-ml-book

[19] Cristina Raluca Gh Popescu and Jarmila Duháček Šebestová. 2024. The impact of artificial intelligence on intellectual capital development: Shifting requirements for professions and processes in the non-profit sector. *Journal of Infrastructure, Policy and Development* 8, 10 (2024). https://doi.org/10.24294/jipd.v8i10.3899

[20] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing.

In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 33–44. https://doi.org/10.1145/3351095.3372873

[21] Warren J von Eschenbach. 2021. Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology* 34, 4 (2021), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

[22] Kate Vredenburgh. 2022. The Right to Explanation. *Journal of Political Philosophy* 30, 2 (2022), 209–229. https://doi.org/10.1111/jopp.12262

[23] Kate Vredenburgh. 2023. AI and Bureaucratic Discretion. *Inquiry* 68, 4 (2023), 1091–1120. https://doi.org/10.1080/0020174X.2023.2261468

A Thematic Network

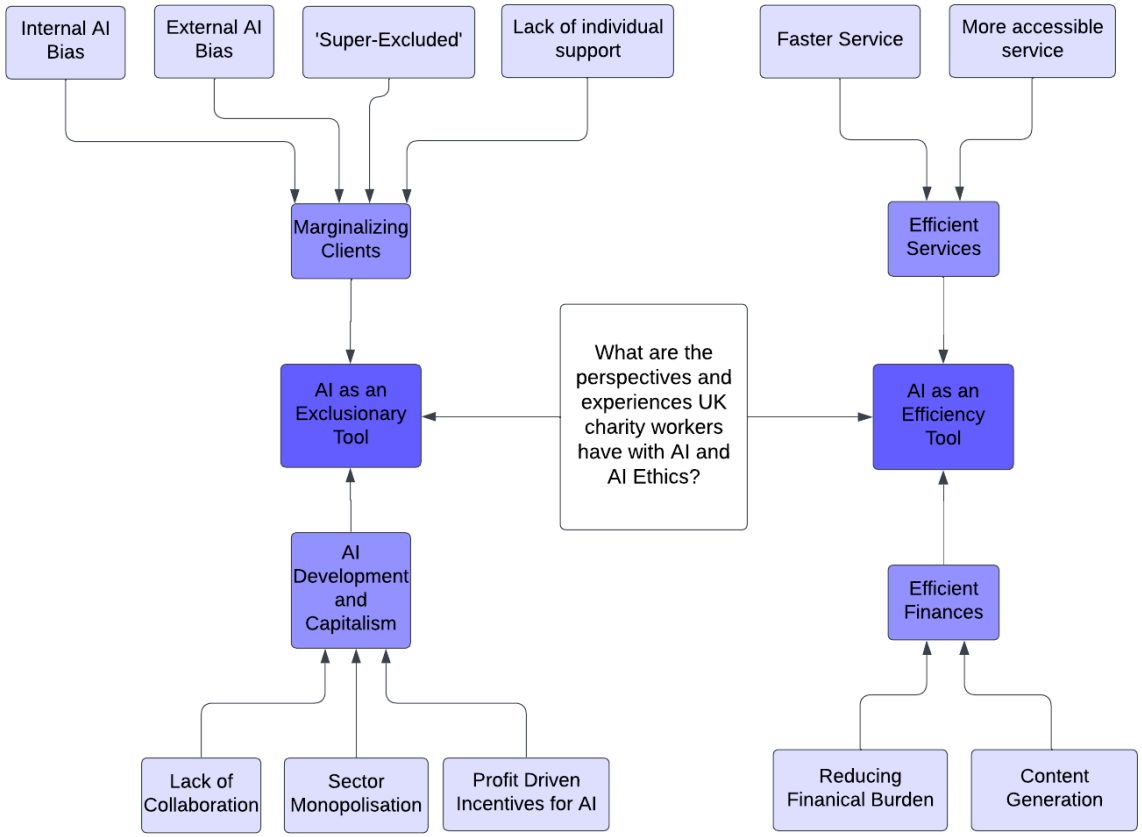


Fig. 1. Full thematic network from analysis of interview transcripts, showing the two global themes, four organizing themes and eleven basic themes or codes.

B Codebook

Table 1. Codebook from the thematic analysis, illustrated by the thematic network in Figure 1

Global Theme	*Organizing Theme*	*Basic Theme*	*Description*	*Example*
AI as an Exclusionary Tool	Marginalizing Clients	Internal AI bias	There is the risk of AI algorithms used by charities to have bias and discriminate against already minoritized clients, leading to worse services and outcomes	"It may be our algorithms that are making the inappropriate decisions if we're not careful" "we're very conscious that the risk due to underlying bias .. is that people who are already kind of disadvantaged, so minoritized clients, will be disadvantaged even more."
		External AI bias	AI used by other organisations have consequences for charity clients, and the charities then must deal with the repercussions	"... they can have about 100% turnover rate because they can challenge the decision by understanding how those decisions have been made. They have very low digital maturity, they can't understand decisions that are made algorithmically. In that kind of predictive future, you see a 100% turnover rate get down to 0." " [we saw] clients of color were being charged more for their vehicle insurance as a result of algorithmic decision making ... I think we're going to be seeing this increasingly in the coming years, and particularly in that consumer space, where people will be experiencing difficult circumstances as a result of algorithmic decision making."

Table 1. Codebook from the thematic analysis, illustrated by the thematic network in Figure 1

Global Theme	*Organizing Theme*	*Basic Theme*	*Description*	*Example*
AI as an Exclusionary Tool	Marginalizing Clients	Super-excluded	Digitally excluded people in society could become even more excluded	"The super-excluded ... I think those people are at risk of just falling off the edge. So there'll be a much smaller group, but there will be much more excluded." "A chatbot would not be suitable for 20% of our clients... The risk there is for that like cohort of people they will become more excluded."
		Lack of individual support	When there is discrimination caused by algorithms, there isn't a mechanism to identify and deal with it immediately; patterns only emerge after time	"You've got charities who are trying to campaign and try to influence change, not being able to, so investigative journalists and data scientists being able to take on that role... It shifts from supporting individuals who have had a negative decision, to supporting cohorts when there's enough data to spot patterns... So what we're really seeing is a movement from individual and community to bigger cohorts." "we haven't got the mechanism to say ... to quickly correct ... If the person who is like delivering the decision doesn't understand how its being made, or because there's so many inputs, no one really understands how it's been made. Then the time lag from decision through to consequence... that's a real kind of challenging thing."

Table 1. Codebook from the thematic analysis, illustrated by the thematic network in Figure 1

Global Theme	*Organizing Theme*	*Basic Theme*	*Description*	*Example*
AI as an Exclusionary Tool	AI Development and Capitalism	Sector monopolization	The fear that larger charities with better tech capabilities will swallow up other charities that do essential work	"There's a risk from the super large... I can see a similar risk in the charity sector, not intentional, where the very large charities take over as digital begins to accelerate. They have very large, capable teams and they're comfortable around tech, they're able to invest in it. If they then become super dominant, it will cut the throat of other charities." "One of their main fears about AI is predatory competitors who make better use of these technologies... there's a lot of fear about other organizations having like a technical competitive advantage."
		Lack of collaboration	Due to competition for funding, charities do not collaborate on AI strategy	"... they're building AI for themselves. That's the whole point about the sector, it's all about me." "[redacted organisation] has a bit of a reputation for not collaborating very much... It's like really odd, because it seems like really at odds with what we're supposed to be doing doesn't it? But I do think its true and it's about competition for funding."

Table 1. Codebook from the thematic analysis, illustrated by the thematic network in Figure 1

Global Theme	*Organizing Theme*	*Basic Theme*	*Description*	*Example*
AI as an Exclusionary Tool	AI Development and Capitalism	Profit driven incentives for AI	The development and funding given for AI is tied very closely to business in such a way that it becomes much less applicable to charities	“I most align with ... community driven governance, where the impacts of AI are equally shared across society, rather than driven specifically by a business model" “The thing that leaves me still unconvinced is how tied the growth of AI is to extractive business models." “These pressures come from government funders looking to or making assumptions about how services can be more cost effective without thinking about how they can be more inclusive. I think people also have incorrect mental models around who is digitally excluded. I think a lot of people will probably come into it thinking it was basically older generations. But here poverty comes into it a great deal. Data poverty is a real thing. "

Table 1. Codebook from the thematic analysis, illustrated by the thematic network in Figure 1

Global Theme	*Organizing Theme*	*Basic Theme*	*Description*	*Example*
AI as an Efficiency Tool	Efficient Services	Faster service	AI can make services faster	There are loads of positives around efficiencies, and enabling people to get immediate decisions." "From a generative AI point of view, we've started work to see how we can make use of these technologies in improving the provision of advice... we have more demand for advice than we can supply. So it's important to us to see how we can introduce efficiencies in that workflow."
		More accessible service	AI can make services more accessible	"... in an AI world accessibility should become much better. But for [the digitally excluded] I think there's a role for the smaller charities ... we should be recognizing that as we begin to design the systems, we should be thinking about the people for whom that will never work." "I also think there's an opportunity if you approach it in the right way to make services more inclusive... Using AI, but with an intermediary who works with a client who would otherwise be excluded from its benefits. "

Table 1. Codebook from the thematic analysis, illustrated by the thematic network in Figure 1

Global Theme	*Organizing Theme*	*Basic Theme*	*Description*	*Example*
AI as an Efficiency Tool	Efficient Finances	Reducing Financial Burden	AI can reduce the financial burden on charities	“The sector is in desperate straits. It never recovered from 2008... There’s a huge funding gap, there are increasing numbers of charities going to the wall ... AI will do two things. It will augment your capabilities, and it will reduce admin workload ... It can’t fix the funding gap, but it can depressurize it." “The benefit of those efficiencies is that they have more time to spend with their friends, families and communities ... we can benefit from all these efficiencies as a society, rather than just those who are able to create businesses and make more efficient businesses"
		Content Generation	AI can produce content, eradicating the need for expensive content generators	“one of the areas of change is the ability for generative AI to interpret the law into accessible text" “Portraying vulnerable beneficiaries has always been a real challenge for us... we were dealing with abused women, alcohol, drugs and you have to be careful about how you use imagery, but actually with AI, you could just create it... But you need to be absolutely transparent about it."

C Topic Guide

Interviews were semi-structured using the following topic guide, with a time guide of 60 minutes. Some questions were asked out of order and not all questions were asked to all participants, although each topic was addressed.

Opening

(1) Introductions
(2) Explain the purpose of the research
(3) Explain confidentiality of the interview
(4) Confirm consent and ask to record
(5) Opening question: Could you tell me a bit about your charity and the work you do?

Interactions with Algorithms and AI
Objective: To understand the interviewees' practical experiences with AI

(6) Could you tell me about the type of interactions you typically have with algorithmic decision making or AI?
(7) What effect does algorithmic decisions making have on your ability for you/charities you work with to do your work effectively?
(8) What level of access do you typically have to external algorithms?
(9) What are your current strategies to responding to decisions made by algorithms rather than people?

Opinions on Algorithms and AI
Objective: To understand the interviewees' current opinions, positive and negative, on AI, and gauge current their current view on AI ethics.

(9) What do you think are the pros and cons arising from algorithmic decision making?
(10) When I say, 'ethical AI', what comes to mind to you?
(11) Do you have any examples of an ethical issue regarding AI you would like to share from your work?

Opinions on AI Ethics Solutions
Objective: Understand the interviewees' opinions on AI Ethics solutions. This is a user-study portion of the interview.
Show a document (see Appendix D) containing an example with explainability and fairness methods. This will be a simulated scenario with the aim being to show the participants that these methods for explaining algorithmic outputs exist and how they could look in a scenario, for example a hiring algorithm that decides if someone gets a job. If they have given an example from their work in the previous questions, tie that in. Explain the data and graphs and ask if they have any questions before proceeding.

(12) These are some methods that researchers have worked to develop to understand the decisions made by algorithms. Have you come across any of these before. What do you think of them?
(13) Which of these do you find most helpful and why?
(14) How could these types of methods help your ability to respond to decisions made by algorithms that you deal with?

Emotional Reactions and Feelings about AI

Objective: Understand what kind of emotions and feelings the interviewee has regarding the topics discussed. If they don't respond with feelings, re-ask the question.

(15) How do you feel about the use of algorithmic decision making and AI?
(16) How does AI ethics solutions affect how you feel about AI in charity work?

Concluding

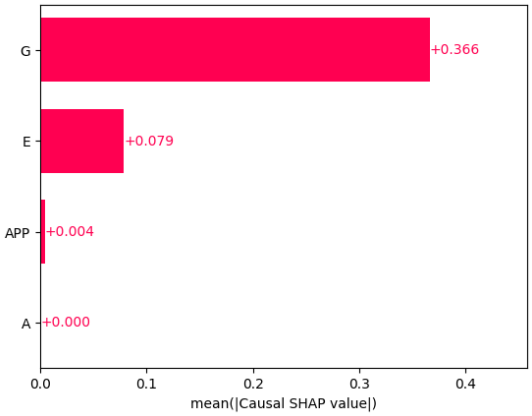
(17) Are there other ethical issues you think academics and researchers should focus on that would help you deal problems arising from AI?
(18) Is there anything else you would like to add?
(19) In the case of any clarification I might need for my research, is it okay to reach out to you with any follow-up questions?
(20) Thank for time and conclude interview

D Interview Supporting Document

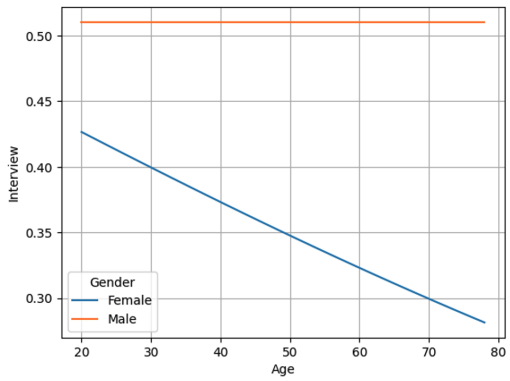
The example here is entirely synthetic, with simulated data and an imagined context. However, it is based on real life scenarios where hiring algorithms have been discriminatory. Imagine that there is a hiring algorithm being used by a company or organisation, and you want to check if the algorithm discriminates against the gender or age of the individual applicant. The algorithm uses the following variables, relevant experience, likelihood of applying, age and gender to decide whether the applicant should be invited for interview. Below are some examples of model-agnostic explanation (or XAI) methods for this scenario on different algorithms. This is not extensive – there are several other methods, these are a sample of some of the most popular.

Speaker Note: Spend time describing the plots and underlying method at a high level (minimal technical detail).

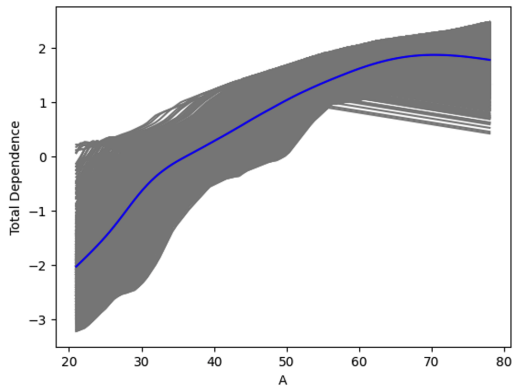
SHAP Plot [15]



Partial Dependence Plot [8]



Total (Causal) Dependence Plot [14]



Accumulated Local Effects Plot [2]

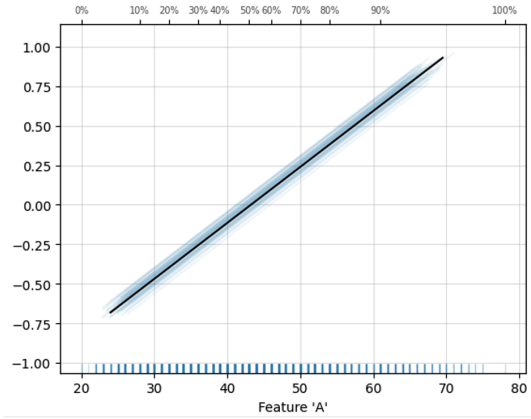


Fig. 2. Figures shown to interviewees to explain interpretability methods for predictive AI.

Popular fairness measures include equal opportunity, demographic parity, treatment equality, among others, which all involve various uses of the true and false positive rates. In the above example, an equal opportunity score of 1 would indicate high unfairness towards female applicants, whereas a score of 0 would indicate fairness between groups. Importantly, it is often the case where one measure shows unfairness, but another does not.

More resources:

- Interpretable Machine Learning Online Book by Christoph Molnar [18]
- Fair Machine Learning Online Book by Solon Barocas, Moritz Hardt, Arvind Narayanan [5]